

# **TDK-dolgozat**

Dénes Botond

# Vízközeli hulladéklerakók megbízható detektálása multispektrális műholdfelvételek segítségével

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

INFORMATIKAI KAR

PROGRAMOZÁSELMÉLET ÉS SZOFTVERTECHNOLÓGIAI TANSZÉK



*Szerző:*

Dénes Botond

programtervező informatikus MSc

2. évfolyam

*Témavezető:*

Cserép Máté

egyetemi tanársegéd

Budapest, 2024

# Tartalomjegyzék

<b>1. Bevezetés</b>	<b>2</b>
1.0.1. A kutatólabor már meglevő eredményei . . . . .	2
<b>2. Elemzés és tervezés</b>	<b>4</b>
2.0.1. Kutatási cél . . . . .	4
<b>3. Betanítás</b>	<b>5</b>
3.0.1. Tanítóadatok . . . . .	5
3.0.2. Tanítási paraméterek . . . . .	6
<b>4. Megvalósítás és alkalmazás</b>	<b>8</b>
4.0.1. A meglevő alkalmazások bővítése . . . . .	8
4.0.2. A Tiszta-Tisza alkalmazás . . . . .	8
4.0.3. Közös metszet . . . . .	9
<b>5. Összefoglalás és eredmények</b>	<b>10</b>
<b>Köszönetnyilvánítás</b>	<b>11</b>
<b>A. Szimulációs eredmények</b>	<b>12</b>
<b>Irodalomjegyzék</b>	<b>13</b>
<b>Ábrajegyzék</b>	<b>14</b>
<b>Táblázatjegyzék</b>	<b>15</b>
<b>Algoritmusjegyzék</b>	<b>16</b>
<b>Forráskódjegyzék</b>	<b>17</b>

# 1. fejezet

## Bevezetés

A hulladékszennyezés komoly problémát jelent a természet számára [1]. Emiatt számos szervezet mozdul abba az irányba, hogy tisztábbá tegye a bolygónkat. Egy ilyen szervezet a PET Kupa, akik elsősorban folyómenti hulladékgyűjtéssel foglalkoznak elsősorban Magyarországon, de figyelmük kiterjed a szomszédos országokra is. Az egyik nagy kihívás a szemétgyűjtésben ezeknek a területeknek megtalálása. Sok emberi és pénzügyi erőforrást igényel a hulladéklerakók megtalálása a folyók mentén, hiszen járművekkel valakinek végig kell haladnia egy hosszabb területen, csak azért, hogy felmérje, hogy hol van hulladék. Ehhez jelentős mennyiségű üzemanyagot kell elhasználni. Emiatt hatékonyabb eszközökre van szükségünk, hogy ezt a folyamatot felgyorsítsuk. Ennek fényében a PET Kupa felkereste az egyetemünket azzal a kéréssel, hogy olyan eszközöket fejlesszünk le, melyek automatikusan képesek lesznek hulladékot detektálni a folyók mentén.

Ide esetleg egy hivatkozás?

A dolgozatomban bemutatok egy Random Forest modell-t, mely a kutatólaborban már lefejlesztett modellre épül [2]. A bemutatott modell javít a korábbi modell problémáin, illetve nagyobb megbízhatósággal találja meg a hulladékot a folyókon és a folyók mentén. A modell eredményei integrálásra kerülnek a Tiszta Tisza webalkalmazásba, ahol több napon keresztül történő detektálás eredménye lesz összesítve és megjelenítve a felhasználók számára.

Hivatkozni a weboldalra

### 1.0.1. A kutatólabor már meglevő eredményei

Megemlíteni, hogy kibővíttem a szerveralkalmazást

A térinformikai kutatólaborban már fejlesztésre került egy szerveralkalmazás, mely minden nap a Planet-ről letölti a legfrissebb felvételeket a vizsgált területekről, és lefuttatja ezeken a képeken az akkori modellt. Ezen felül készült egy webalkalmazás is,

ami erről a szerverről letölti az eredményeket, és megjeleníti ezeket, összehasonlításra. A kutatólabor rendelkezik egy asztali alkalmazással is, mellyel hatékonyan elő lehet állítani tanítóadatokat. A kutatásom elősegítéséhez ezeket az alkalmazásokat használtam, illetve bővítettem a 4.0.1 fejezetben leírtak szerint.

## 2. fejezet

# Elemzés és tervezés

### 2.0.1. Kutatási cél

A cél az, hogy a kutatás során szerzett modell megbízhatóan detektáljon hulladéklerakókat. Ehhez a false positive arányok minél kisebbek kell legyenek, míg a true positive arányok minél nagyobbak. Ugyanakkor nem jelent ugyanakkora problémát egy false negative, mint egy false positive, mivel a false positive eredmények fölöslegesen rossz irányba küldhetik a folyómentő csapatot. A kutatólabor 2023-as cikkjében bemutatott modell (továbbiakban meglevő modell) egyik problémája a nagy false positive arányok voltak. A modell a pusztazámori hulladéklerakóról, illetve a kiskörei víztárolóról szerzett adatokkal volt betanítva. Ezért érdemes egy nagyobb adathalmazzal betanítani a modellt.

Az új Random Forest modell a PlanetScope műholdakra lesz specializálva, azon belül is a legújabb PSB.SD szenzorokra. A modell számára elérhető lesz a Vörös, Kék, Zöld, és a közeli infravörös (NIR) sáv. A PlanetScope műholdak körülbelül 3 méter/pixel felbontással rendelkeznek [3, 4].

## 3. fejezet

# Betanítás

### 3.0.1. Tanítóadatok

A betanításhoz 29 romániai hulladéklerakó és közvetlen környezete került a tanítóadatok közé, illetve a Kiskörei víztároló is. A romániai hulladéklerakókat egy helyi weboldalon lehet megtalálni, a hozzájuk tartozó koordinátákkal együtt [5]. Az ott bemutatott 46 hulladéklerakó közül 29 volt alkalmas tanításra: sok hulladéklerakó be lett tömve, vagy föld alatt működik. Minden hulladéklerakóhoz tartozik egy-egy nyári+tavaszi, téli és őszi multispektrális műholdkép, melyeket kézzel annotáltam. A nyári és tavaszi képeket azért vontam egybe, mivel ezek hulladékdetektálás szempontjából hasonló adatokat eredményeztek. A tanítóadatok pixelenként vannak előállítva, így a végső adathalmaz 27 millió tanítóadatból (pixelből) áll. Minden pixelhez hozzá van rendelve a vörös, kék, zöld, közeli infravörös sáv, illetve a "PI", "NDWI", "NDVI", "RNDVI", "SR" indexek. Ezen felül minden pixel címkézve van a 3.1 táblázatban leírtak szerint.

Címke azonosító	Címke neve	Címke magyarázat
100	Hulladék	Azon területek, melyeken hulladékot találtunk.
200	Víz	olyan területek, melyeken kizárólag vizet találtunk, általában folyók.
300	Legelők/Erdők	Zöld övezetből álló vad területek. Ezek lehetnek fák lombjai vagy füves zónák.
400	Mezők	Olyan földes területek, melyek meg vannak művelve, illetve ahol mezőgazdasági növények találhatók, például gabonafélék.
500	Ismeretlen	Olyan területek, melyek a korábbi kategóriákba nem sorolhatók bele. Ilyenek az épületek, aszfaltozott utak, háztetők, mezei utak.

3.1. táblázat. A tanítóadatok címkéi

### 3.0.2. Tanítási paraméterek

A nagy adathalmaz miatt a Random Forest modell is nagyon nagy lesz (körülbelül 14GB), ami egy nehezen kezelhető méret, így érdemes módosítani a modell paraméterein, hogy ez kisebb méretű legyen. A legjobb eredményeket azzal értem el, hogy a Random Forest fák méretét 20 mélységűre limitáltam. Ennek köszönhetően a modellek méretét 2GB-ra tudtam csökkenteni, és a ?? ábrából látható, hogy a csökkentett modell is hasonlóan teljesít a nagy modellhez képest.

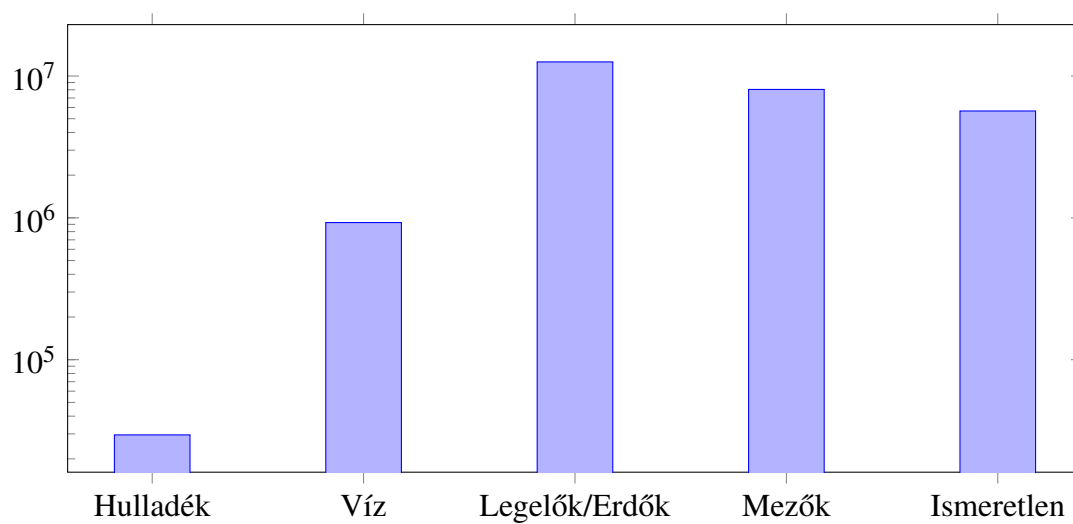
Továbbiakban felmerült az a probléma is, hogy a tanítóadatok nagyon aránytalanok voltak: A 3.1 ábrából látható, hogy kevesebb adattal rendelkezünk hulladékról, mint az összes többi adatról, így a modell nagyon sok false-negatívot termelt. Ennek korrigálására súlyokat alkalmaztam a tanítóadatokra. A súlyok kiszámolásához az összes címkére a 3.1 képletet használtam.

$$\text{címke súly} = \frac{\text{adathalmaz mérete}}{\text{címke darabszáma}} \quad (3.1)$$

táblázat a fák méretéről, a modellek méretéről és a különböző mélységekről

ábra készítés ide





3.1. ábra. Az adatok közötti aránytalanság, logaritmikus skálázással

## 4. fejezet

# Megvalósítás és alkalmazás

### 4.0.1. A meglevő alkalmazások bővítése

A meglevő asztali alkalmazás alkalmas volt a tanítóadatok hatékony előállítására, de utólag nem lehetett visszanézni, hogy adott műholdfelvételhez milyen tanítóadatok tartoznak, illetve azt sem, hogy az adott tanítóadat hol volt mintavételezve. Az alkalmazás eredetileg egy CSV fájlba eltárolta az összes pixel spektrális értékeit és indexeit, és ezt lehetett használni tanításra. Ennek az volt a hátránya, hogy nehéz volt áttekinteni illetve kiegészíteni az adatokat. Ezért az asztali alkalmazást kiegészítettem ezzel a funkcionalitással, a tanítóadatok előállítása elmentésekor az alkalmazás létrehoz egy külön raszteres réteget is külön minden műholdfelvételhez, melyen látható, hogy mely területek voltak hozzáadva a tanítóadatok közé, így tetszőleges módon előállítható/ellenőrizhető a tanítóhalmaz.

A szerveralkalmazás és webalkalmazás is bővítésre került: a szerveralkalmazás mostmár több modellt is le tud futtatni a letöltött műholdfelvételeken és ezeket külön tárolja. A webalkalmazás mostmár képes letölteni külön ezeket az eredményeket és több hulladékmaszkoló módszer eredményét is meg tudja jeleníteni, ennek köszönhetően ezeket egymással össze lehet könnyen hasonlítani valós tesztadatokon.

### 4.0.2. A Tiszta-Tisza alkalmazás

melyik link  
kerüljön ide?

A Tiszta-Tisza webalkalmazás a PET Kupa által használt webalkalmazás, melynek az a célja, hogy egy olyan felületet biztosítson, ahol meglehet tekinteni a jelenleg ismert folyómentén található hulladéklerakókat, illetve akár a regisztrált felhasználók is be tudnak jelenteni ilyet. A PET Kupa megbízta az egyetemet azzal a feladattal, hogy

ezt továbbfejlessze, és a feladatok közé tartozott az is, hogy a Random Forest modell eredményeit integráljuk ebbe az alkalmazásba. Ezt a feladatot én vállaltam el.

Tekintve arra, hogy a Tisza-Tisza térképén pontok vannak megjelenítve, a modell által detektált területeket is pontokkal jelöljük. Ehhez egy nagyobb terület közepére helyezünk el egy pontot. Előfordulhat olyan is, hogy a modell olyan képeket klasszifikál, melyek el vannak torzítva (például magas páratartalom miatt). Ilyenkor a false-positive-ok aránya lényegesen megnő. Ennek korrigálására a Tisza-Tisza alkalmazásban a legutolsó három detektálást (legfeljebb 1 hónap különbséggel) veszem figyelembe és két kép közös metszetével döntöm el, hogy milyen területek kerülnek fel a térképre. A lépéseket a 4.0.3 fejezetben részletezem.

### 4.0.3. Közös metszet

A már meglevő szervertől poligonok formájában, GeoJSON-ben [6] lehet lekérni az adott napon detektált hulladékterületeket. Így érdemes poligonok metszetében kigondolni a többségi szavazást. Jelöljük  $BUF(P,n)$ -vel egy multipoligon puffert, ahol

$$P \in \mathbb{P}$$

egy multipoligon, és  $n$  egy egész szám. Ekkor a többségi szavazást a 4.1 képlet szerint lehet alkalmazni. Ezután az elég nagy poligonok egy-egy belső pontját megválasztva megtudjuk jelölni a hulladéklerakókat.

$$Eredmény\ multipoligon = \bigcup_{P_1 \in \mathbb{P}} \bigcup_{P_2 \in \mathbb{P}} BUF(BUF(P_1, n) \cap BUF(P_2, n), -n) \quad (4.1)$$

használt  
képleteket  
mások is  
ismernek  
fórumokban,  
de hivatalos  
forrással nem  
találkoztam

## **5. fejezet**

### **Összefoglalás és eredmények**

# Köszönetnyilvánítás

Amennyiben a TDK projektet pénzügyi támogatást kapott egy projektből vagy az egyetemtől, jellemzően kötelező feltüntetni a dolgozatban is. A dolgozat elkészítéséhez segítséget nyújtó oktatók, hallgatótársak, kollégák felé is nyilvánítható külön köszönet.

## **A. függelék**

### **Szimulációs eredmények**

# Irodalomjegyzék

- [1] M.G. Kibria, N.I. Masuk és R. et al. Safayet. „Plastic Waste: Challenges and Opportunities to Mitigate Pollution and Effective Management”. *International Journal of Environmental Research* 17.20 (2023. jan.). ISSN: 2008-2034. URL: <https://doi.org/10.1007/s41742-023-00507-z>.
- [2] Dávid Magyar és tsai. „Waste Detection and Change Analysis based on Multispectral Satellite Imagery”. 2023. jan. DOI: 10.48550/arXiv.2303.14521.
- [3] Planet.com. URL: <https://developers.planet.com/docs/apis/data/sensors/> (elérés dátuma 2024. 04. 03.).
- [4] Planet.com. URL: <https://developers.planet.com/docs/apis/data/sensors/> (elérés dátuma 2024. 04. 03.).
- [5] InfoCons.ro. URL: <https://fiiunexemplu.ro/in-romania-exista-46-depozite-de-deseuri-gropi-de-gunoi/> (elérés dátuma 2024. 04. 02.).
- [6] H. Butler és tsai. *The GeoJSON Format*. RFC 7946. 2016. aug. DOI: 10.17487/RFC7946. URL: <https://www.rfc-editor.org/info/rfc7946>.

# Ábrák jegyzéke

3.1. Az adatok közötti aránytalanság, logaritmikus skálázással . . . . .	7
--	---



# Táblázatok jegyzéke

3.1. A tanítóadatok címkéi . . . . .	6
--------------------------------------	---

# **Algoritmusjegyzék**

## **Forráskódjegyzék**